Dialogue Evaluation 2019

We are pleased to announce two shared tasks for people interested in coreference and anaphora resolution and machine learning in general. Both shared tasks will be held in 2019 in the framework of the [Dialogue](#) conference.

Important dates:

Registration closes Jan 25th 2019
Release of the Training Data Jan 26th 2019
Release of the Test Data Feb 15th 2019
Systems Submissions due Feb 22th 2019
Final results from organizers Mar 5th 2019

1. Shared Task for Anaphora and Coreference Resolution for Russian (AnCorR)

High-quality coreference resolution plays an important role in many NLP applications. However, developing a coreference resolver for a new language requires extensive world knowledge as well as annotated resources, which are usually expensive to create.

The aim of anaphora and coreference resolution components of an NLP system is to find all mentions in the text that refer to the same real-world entity. The first such evaluation for Russian was organized in 2014 ([RU-EVAL 2014](#)). The latest Shared Tasks for multilingual coreference are, e.g., [CORBON 2017](#) (where Russian was one of the concerned languages) and [CoNLL-2012](#).

Task Description
The shared task is divided into coreference and anaphora resolution tasks.

In coreference resolution task, The training set has two layers. This allows not only to train a system to determine whether two mentions are coreferential (coreference chains layer) but also to localize the boundaries of the mentions (mentions layer). In coreference chains layer, for each mention included in a chain of length more than one, there is a line describing it in the following format: Mention ID→Mention Offset→Mention Length→Chain ID. In mentions layer, for *each* mention in a text, there is a line describing it in the following format: Mention ID→Mention Offset→Mention Length. Mention IDs are sorted in order of appearance in the text. Mentions with equal IDs in both layers have equal offsets and lengths.

Anaphora resolution task, the training set consists of anaphoric pronouns and their antecedents.

Examples (data)
Петя$_1$ попросил Васю помочь с домашним заданием. Мальчик$_1$ плохо разобрался с темой.
Peter$_1$ asked John to come home because the boy$_1$ was scared of being alone.
Мама мыла раму$_{ante}$. Теперь она$_{ana}$ чистая.

Theresa Mary May$_{ante}$ is a British politician serving as the current Prime Minister of the United Kingdom. She$_{ana}$ identifies herself as a one-nation conservative.

Participants may choose one of the tracks of try them all. Systems' performance will be measured with averaged metric of MUC, B-cube and CEAF (Cai, Strube, 2010)*. The organizers will provide a scoring system and the output of some baseline systems.

*Jie Cai and Michael Strube, 2010, Evaluation Metrics For End-to-End Coreference Resolution Systems

Link for the registration
[Anaphora Resolution](#)

2. AGRR: Automatic Gapping Resolution for Russian

Gapping is the most common type of ellipsis, concerning such examples as
"Ей он рассказывает одно, а нам — совершенно другое",
"Кто любит арбуз, а кто - свиной хрящик",
"Дайте мне две пятерки, а я вам десятку"

Motivation

The aim of this task is to challenge non-trivial linguistic phenomenon, gapping, that occurs in coordinated structures and elides a repeated predicate, typically from the second clause. Besides the adversity of the construction itself, the phenomenon is naturally rare, which results in lack of training data. During the last two years Gapping has received considerable attention ([S Schuster, M Lamm, CD Manning 2017](#); [K Droganova, D Zeman 2017](#); [K Droganova et al 2018](#); [S Schuster, J Nivre, CD Manning 2018](#); [Nivre et al 2018](#)). Unfortunately, research was mainly held on insufficient data not exceeding several hundreds of sentences so far.
This campaign is a pilot event for gapping resolution task for Russian held for the first time.

Examples (data)

Participants will be provided with a corpus of several thousands of examples coming from texts of different genres, such as news, fiction, and science. Each sentence will be annotated as follows: two remnants R1 and R2, their correlates in the antecedent clause cR1 and cR2, the position of the elided predicate V and the head of the correspondent predicate cV.

(1) Тогда я cV[принял cV] cR1[ее cR1] cR2[за итальянку cR2], а R1[его R1] V[] cR2[за шведа cR2].

(2) *cR1[Иногда cR1] они cV[развиваются cV] cR2[слабо cR2], R1[иногда R1] - V[] R2[очень сильно R2], и тогда они начинают влиять на ход сюжета, а не наоборот.*

Task Description

- **Binary presence-absence classification.** For every sentence decide if there is a gapping construction in it
- **Gap resolution.** Predict the position of the elided predicate and the correspondent predicate in the antecedent clause
- **Full annotation.** In the clause with the gap predict the linear position of the elided predicate and annotate its remnants. In the antecedent clause find the constituencies that correspond the remnants and the predicate that corresponds the gap

Data formats and metrics

The input and output data formats and metrics are specified in the github repository.

AGRR tracks

The following tracks are offered to participants:

1. Closed track – open source track.

- *convenient for research groups and student teams*

Participants are allowed to train their models only on open-access data (open source dictionaries, word embeddings, open parsing systems, etc)

To verify the results, participants should place their code and the model on github, so that it would be publicly available - both for organizers and other teams.

2. Open track - no restriction on data and systems used.

- *recommended for industrial participants, representing their products*

Track participants are allowed to bring any data for learning beyond the data provided and use their own commercial programs. Github sharing is not required.

Participants are welcome to submit their models to both of the tracks under specified constraints.

AGRR repo: https://github.com/dialogue-evaluation/AGRR-2019

Organizers

- Toldova S. Ju, National Research University HSE, Moscow, Russia
- Nedoluzhko A., Charles University, Prague, Czech Republic
- Iomdin L. L., The Institute for Information Transmission Problems, RSUH, Moscow, Russia
- Budnikov E., ABBYY, Moscow, Russia
- Ponomareva M., ABBYY, Moscow, Russia
- Smurov I., ABBYY, Moscow, Russia
- Shavrina T.O., NRU HSE, Sberbank, Moscow, Russia

- Droganova K., Charles University, Prague, Czech Republic
- Bogdanov A., ABBYY, Moscow, Russia

Link for the registration
[Gapping Resolution](#)

Ask your questions here
dialogueeval2019@gmail.com